DOCUMENT RESUME

ED 294 276 CS 506 143

AUTHOR Nicolai, Michael T.
TITLE The Professional and the Lay Judge: A Comparison of Competitive Rankings in Forensics Tournaments.
PUB DATE Nov 87
NOTE 12p.; Paper presented at the Annual Meeting of the Speech Communication Association (73rd, Boston, MA, November 5-8, 1987).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Communication Research; Comparative Analysis; Debate; Evaluation Criteria; *Interrater Reliability; *Judges; Lay People; *Persuasive Discourse; Public Speaking; Secondary Education; Speech Communication
IDENTIFIERS *Debate Tournaments; Evaluation Research; Evaluation Standards

ABSTRACT

To determine if there is a distinction between the forensics community's idea of quality and that of the general population, tournament rankings of forensics judges and those of a lay audience were compared. Undergraduate students enrolled in a variety of speech related courses were asked to attend rounds of competition at a midwest collegiate forensics tournament and fill out a ranking sheet (similar to a master ballot used by the regular assigned judges) on the performers. The comparative rankings generated 921 ranked pairs. Results showed that there does not appear to be a significant similarity in the direct rankings or adjacent rankings of professional and lay judges. The data also implies that even within the ranks of forensics judges a wide range of ranks is common and a wide spectrum of evaluation is likely. (Eight tables of data are included.) (MS)

THE PROFESSIONAL AND THE LAY JUDGE: A COMPARISON OF COMPETITIVE RANKINGS IN FORENSICS TOURNAMENTS

By

Michael T. Nicolai

University of Wisconsin-Stout

Presented at the Speech Communication Association National Convention, November 7th, 1987

When watching an Olympic figure skater or a gymnast perform there is often a sense of awe concerning the skills being demonstrated. When results are announced it isn't uncommon for audiences to wonder how a judge can distinguish among the performances. Television commentators will offer insight into the nuances of a performance pinpointing technicalities which make or break a routine. The average audience member may marvel at the pickiness, or the highly technical criteria; reactions such as "I liked her the best, regardless of what the judges say" are not uncommon. Forensics competitors perform in a similar environment. There are specific written rules which govern an event, such as time limits, and there are unwritten conventions which have become criteria for evaluation, such as the type of script books for interpretive events. In addition there are regional attitudes which become critical factors in judging, such as the degree of movement or acting in a duo. All of these elements have developed into a highly specialized set of rules which govern the preparation, performance and evaluation/judging of a forensics event. One possible outcome of this is a situation where the forensics community's ideas of quality become dramatically different from the general population. It seems that it would be useful to attempt to determine if such a distinction exists. One possible way to study this is to compare the tournament rankings of forensics judges to those of a lay audience. Two hypothesis were suggested to test the condition.

H 1 "There is a significant similarity in the <u>direct</u> rankings of forensics performances by forensics judges and lay judges."

H 2 "There is a significant similarity in the <u>adjacent</u> rankings of forensics performances by forensics judges and lay judges."

Two terms may require clarification. A "direct" ranking is a pair of identical rankings, e.g. each judge ranks a contestant first. An "adjacent" ranking exists when judges rank within one place of each other, e.g. one judge ranks a contestant second and the other judge ranks that contestant first or third. The rationale for considering the second hypothesis will be explained in the results section.

METHODOLOGY

Undergraduate students enrolled in a variety of speech related courses were asked to attend rounds of competition at a midwest collegiate forensics tournament and fill out a ranking sheet on the performers. This sheet was similar to a master ballot used by the regular assigned judges. The student/lay rankers were randomly assigned to events and sections. In some instances more than one lay judge ranked the same set of performers. The forensics judges and lay judges were instructed not to collaborate with each other.

For purposes of this initial study the comparative rankings were compiled for two interpretive events, (poetry and drama interpretation) and two original events, (informative speaking and persuasive speaking). This generated 921 ranked pairs including 114 informative rankings, 270 persuasive rankings, 216 drama interpretation rankings and 321 poetry interpretation rankings.

The paired rankings for each event were subjected to chi-square cross tabulations. An additional cross tabulation was generated combining the ranked pairs for all the events. This procedure provided data concerning the frequency of direct agreement between the forensics judge and the lay judge.

In order to test the second hypothesis the ranks were analyzed by comparing the frequency of adjacent rankings using the following table:

TABLE 1

Forensics Judge Rank	Lay Judge Adjacent Rank
1st	1st and 2nd
2nd	1st, 2nd, 3rd
3rd	2nd, 3rd, 4th
4th	3rd, 4th, 5th
5th	4th, 5th, 6th

Results

The results of the study related to the first hypothesis are summarized in tables 2-6. The data gathered supports the null hypothesis; there does not appear to be a significant similarity in the direct rankings of forensics and lay judges. The result of the clustering of adjacent rankings is summarized in tables 7-8. This data

suggests that some similarity exists but the significance of this similarity is open to interpretation. The specific results and statistical analysis are as follows:

Cross tabulation of rankings of forensics judge by lay judge

TABLE 2
ALL EVENTS
Lay Judges

	Rank	1	2	3	4	5
Forensics Judge	1	11.2%*	22.9%	15.3%	13.5%	7.1%
		40.2%	27.8%	15.2%	13.9%	5.0%
	2	20.5%	29.8%	14.9%	18.7%	11.1%
		20.1%	29.8%	19.9%	19.1%	7.9%
	3	18.9%	20.1%	26.0%	13.6%	21.3%
		18.1%	19.9%	25.7%	13.9%	15.0%
	4	12.1%	15.3%	19.4%	32.1%	20.6%
		12.1%	15.2%	19.3%	33.3%	11.6%
	5	6.6%	8.7%	14.1%	13.3%	57.3%
		9.2%	12.3%	19.9%	19.1%	57.5%

*The top number is the row percentage and the second number is the column percentage.

With a five by five matrix there are sixteen degrees of freedom resulting in a critical value of 39.252 which would result in a probability level of .001. The chi-square value for the comparison of all ranks is 261.98 which is well beyond the .0000 level of significance. As noted at the beginning of this section this strongly supports the null hypothesis indicating that the lay judge and the forensics judge do not directly agree on the comparative value of a forensics performance. In looking at the raw percentages it can be noted that rankings agree 50% of the time only when considering the 5th place ranking. All judges were instructed to rank no lower than 5th regardless of the number of contestants in a round which meant that more than one

competitor could be ranked 5th thus increasing the likelihood of agreement among judges at this rank. It is interesting to note, however, that the next highest percentage of agreement occurred in the 1st place rankings. This may support the notion held by some that it is easy to identify the top and bottom of a round but the middle three ranks represent a grey area where specific distinction of rank is difficult.

The chi-square comparisons for each of the events tends to follow the pattern of the combined data. There were some peculiarities which are noted following each table.

TABLE 3
Cross tabulation of rankings of forensics judge by lay judge
INFORMATIVE SPEAKING

		Lay Judge				
	Rank	1	2	3	4	5
Forensics Judge	1	38.1%*	28.6%	14.3%	9.5%	9.5%
		38.1%	27.3%	15.0%	9.5%	6.7%
	2	28.6%	38.1%	9.5%	4.8%	19.0%
		28.6%	36.4%	10.0%	4.8%	13.3%
	3	4.8%	14.3%	38.1%	23.8%	19.0%
		4.8%	13.6%	40.0%	23.8%	13.3%
	4	14.3%	9.5%	9.5%	38.1%	28.6%
		14.3%	9.1%	10.0%	38.1%	20.0%
	5	10.0%	10.0%	16.7%	16.7%	46.7%
		14.3%	13.6%	25.0%	23.8%	46.7%

*The top number is the row percentage, bottom number is the column percentage.

The chi-square for this event was 38.91 which reflected is a probability level of .0011. The analysis of this event should be tempered by an understanding that the smaller sample size (114) resulted in 11 cells with an event frequency of less than

five which does affect the final significance assessment. The assessment of raw percentages again revealed that in all cases there was less than 50% direct agreement for any given ranking. Also of interest is the consistency of ranking agreement in the 1st through 4th place ranks.

TABLE 4
Cross tabulation of rankings of forensics judge by lay judge
PERSUASIVE SPEAKING

		Lay Judge				
	Rank	1	2	3	4	5
Forensics Judge	1	55.1%*	26.5%	10.2%	8.2%	0
		52.9%	27.7%	10.0%	8.3%	0
	2	8.2%	22.4%	36.7%	20.4%	12.2%
		7.8%	23.4%	36.0%	20.8%	8.1%
	3	16.3%	30.6%	20.4%	14.3%	18.4%
		15.7%	31.9%	20.0%	14.6%	12.2%
	4	20.1%	6.1%	20.4%	30.6%	22.1%
		19.6%	6.4%	20.0%	31.3%	14.9%
	5	2.7%	6.8%	9.5%	16.2%	64.9%
		3.9%	10.6%	14.0%	23.0%	64.9%

*Top number is the row percentage, bottom number is the column percentage.

The chi-square for persuasive speaking ranks was 142.98 which also was significant beyond the .0000 level. These results seem to most dramatically support the notion of ability to distinguish top and bottom ranks with 1st place rank agreements of 55% and 5th place rank agreements of 65% It is also interesting to note that this is the only instance where no lay judges felt a forensics judge's 1st place was the worst performer in the round.

TABLE 5
Cross tabulations of rankings of
forensics judges by lay judges
DRAMA INTERP

Lay Judge

	Rank:	1	2	3	4	5
Forensics Judge	1	48.7%*	20.5%	15.4%	10.3%	5.1%
		48.7%	20.0%	15.4%	10.5%	3.3%
	2	25.0%	45.0%	5.0%	12.5%	12.5%
		25.0%	45.0%	5.1%	13.2%	8.3%
	3	13.2%	13.2%	36.8%	18.4%	18.4%
		12.8%	12.5%	35.9%	18.1%	11.7%
	4	7.7%	10.3%	15.4%	35.9%	30.8%
		7.7%	10.0%	15.4%	36.8%	20.0%
	5	3.3%	8.3%	18.3%	13.3%	56.7%
		5.1%	12.5%	28.2%	21.1%	56.7%

*The top number is the row percentage, bottom is the column percentage.

The chi-square for drama interp was 103.29 with a significant level of .0000. This event appeared to have the highest incidence of direct agreement at each rank with a range of 36 to 57% direct agreement.

TABLE 6 ON NEXT PAGE

TABLE 6
Cross tabulation of rankings of
forensics judges by lay judges
POETRY INTERP

		Lay Judge				
	Rank	1	2	3	1	5
Forensics Judge	1	26.2%* 25.4%	19.7% 19.4%	19.7% 19.4%	21.3% 22.4%	13.1% 10.5%
	2	21.6% 23.0%	23.0% 22.6%	19.7% 19.4%	26.2% 27.6%	6.6% 5.3%
	3	29.5% 28.6%	18.0% 17.7%	19.7% 19.4%	6.6% 6.9%	26.2% 21.1%
	4	8.2% 7.9%	27.9% 27.1%	24.6% 24.2%	29.5% 31.0%	9.8% 7.9%
	5	11.7% 14.3%	10.4% 12.9%	14.3% 17.7%	9.1% 12.1%	54.5% 55.3%

* Top number is the row percentage, bottom is the column percentage.

For the poetry interpretation rank comparisons the chi-square was 80.99, again was a significance level of .0000. The results paralleled the other events in terms of percentage patterns.

When considered individually or collectively it becomes apparent that there is not a high degree of direct agreement between the forensics and lay judge. However, it might be argued that as long as the two are close in their rankings then there may be a case for suggesting some form of agreement. What then becomes necessary is to generate a definition of a "close" ranking. If a performance is evaluated by two judges so that only one rank separates the assessment it might be argued that these judges are "close" in their rankings. To meet this criteria ranks were collapsed as indicated in Table 1 in the previous section. The results are summarized in Table 7.

TABLE 7

ADJACENT RANKINGS

Forensics Judge Rank	Lay Judge Adjacent Rank	Percentage of Occurrence All Events	Informative	Persuasion	Drama	Poetry
1st	1st/2nd	64.1	66.7	69.2	69.2	45.9
2nd	1st/2nd/3rd	70.2	76.2	67.7	75.0	67.3
3rd	2nd/3rd/4th	59.7	76.2	65.3	68.1	44.3
4th	3rd/4th/5th	72.4	76.2	73.4	82.1	63.9
5th	4th/5th	70.6	63.4	81.1	77.7	63.6
Ave. % of occurrence		67.4	71.7	73.8	74.5	57.0

This data does suggest that there is a degree of "closeness" between the rankings of forensics and lay judges. When the results of rankings for all events is collapsed there was an average of 67.4 percent close agreement. However, if considered from the opposite viewpoint it might be argued that even with combined ranks there was still a high percentage of ranks that were not "close". As Table 8 suggests, the range was from 25 percent to 43 percent of the rankings which fell outside a "close" similarity.

TABLE 8

OCCURRENCE OF WIDELY DIVERGENT RANKS

All Events	32.6%
Informative Speaking	28.3%
Drama Interpretation	25.5%
Persuasive Speaking	26.2%
Poetry	43.0%

Depending on the perspective taken this collapsing of ranks may or may not suggest a similarity between forensics and lay judges. However, when considered in light of the earlier chi-square factoring the evidence seems to suggest a rejection of the second hypothesis.

Discussion/Conclusions

The implications of this study are directly related to a number of factors. Initially it must be noted that even within the ranks of forensics judges a wide range of ranks is common. This is something that occurs regularly during final rounds of competition where multiple judges are used. (As a side bar it should be

pointed out that in final rounds the difference between performances is often so small that a judge may feel 1st through 5th ranked performances are near identical, a situation which is not as common in preliminary rounds). Secondly, it can be argued that forensics is an art form and as such a wide spectrum of evaluation is likely.

Yet it must also be noted that this art form is approached quite scientifically by performers, coaches, and judges. It is an art form with many rules. And it is this set of rules that may be the real cause for the dissimilar rankings noted in this study. It seems that a question should be asked about why a lay audience would feel so different about a performance? Could it be that the standards and guidelines employed by the forensics community and reflected in the judging of performers have become more representative of a specialized environment than of society as a whole? And this leads to a most critical question: If this is true is it bad? Is it desirable that the forensic community's standards mirror a lay audience's criteria of what makes a performance "good"? Referring back to the opening analogy, a case can be made that, like the Olympic figure skater, a forensics competition represents the upper echelon of public communicators and therefore should be judged by a set of standards that are above the norm. Such a position is both noble an educationally sound. However, that same zest for the ideal may result in the acceptance of criteria which may or may not be doing justice to the performer. For instance it is not uncommon for a performance to be judged in part on the number of times the material has been done in the past. Material that is "well traveled" is often assessed negatively by forensics judges yet a lay audience would most likely not consider the frequency of performance as a major criteria. As another example forensics judge may negatively evaluate a performer because they open their script during the introduction, yet this wouldn't bother a lay audience. These examples suggest that there may be a question concerning the appropriateness and usefulness of some of the judging criteria used in this activity.

It may well be that the distinctions created are desirable, if so, then results such as those in this study are of little consequence. On the other hand if a major

aspect of forensics is to help student competitors reach a rarified level of communication behavior which is readily adaptable to a general population and setting, then this study may indicate the need for further investigation into the standard used by forensics activists.